Subject: Initial Approach to Testing the Hypothesis on Churn Due to Price Sensitivity

Dear AD ,

I hope you are doing well. I've been reviewing our client's request to test the hypothesis that customer churn in the SME segment is driven by price sensitivity. To effectively test this hypothesis and develop a predictive model for identifying at-risk customers, I have outlined the major steps and data requirements for our approach:

**Formulating the Hypothesis as a Data Science Problem:** Our data science problem can be defined as follows:

Problem: Predict customer churn in the SME segment.

Hypothesis: Churn is influenced by price sensitivity, and offering a 20% discount can be an effective retention strategy.

Goal: Develop a predictive model that identifies customers at risk of churning due to price sensitivity.

**Major Steps:**

**Data Collection:**

Historical data on customer churn: We need records of churn events, including when and why customers left.

Customer profiles: Demographic information, contract details, and historical interactions.

Usage patterns: Data on electricity and gas consumption.

Pricing data: Records of price changes, discounts, and promotions offered to customers.

Communication history: Customer interactions, complaints, and feedback.

**Data Preparation and Feature Engineering:**

Create a comprehensive dataset with relevant fields, including customer ID, demographics, contract type, contract duration, historical price changes, and usage patterns.

Engineer features that capture price sensitivity, such as the frequency of price changes and the magnitude of discounts.

Calculate a churn label based on historical data (e.g., if a customer left within the last 6 months).

**Exploratory Data Analysis (EDA):**

Conduct EDA to understand the distribution of key variables and their relationships with churn.

Identify correlations between price changes, customer demographics, and churn rates.

Visualize the distribution of price changes over time and its impact on churn.

**Model Selection:**

Choose an appropriate predictive model for customer churn. Common models include logistic regression, decision trees, random forests, and gradient boosting.

Given the large dataset, consider models that can handle complex relationships.

**Model Training and Validation:**

Split the data into training and validation sets.

Train the selected model to predict customer churn based on price sensitivity and other relevant features.

Evaluate model performance using appropriate metrics (e.g., accuracy, precision, recall).

**Deployment for Monthly Predictions:**

Develop a pipeline for the predictive model that can handle new data each month.

Implement the model to identify customers at risk of churning due to price sensitivity on the 1st working day of every month.

**Monitoring and Feedback:**

Continuously monitor the model's performance in identifying at-risk customers.

Gather feedback from customer service teams and customers regarding the effectiveness of the 20% discount strategy.

**Reporting and Recommendations:**

Provide regular reports on model performance, including its accuracy, precision, recall, and the number of retained customers.

Make recommendations for adjusting the discount strategy based on the model's output and feedback from the field.

In summary, this approach will allow us to test the hypothesis that churn is driven by price sensitivity and develop a predictive model to identify at-risk customers. It is important to gather and analyze historical data, engineer relevant features, and evaluate the model's performance rigorously.

I look forward to your feedback and any additional insights you may have on this approach. Please let me know if you have any specific data sources or fields in mind that would be particularly valuable for this analysis.

Thank you for your guidance, and I am excited to work on this project.

Best regards,

Ajish Daniel